

**CENTER FOR DRUG EVALUATION AND
RESEARCH**

APPLICATION NUMBER:

21-303

STATISTICAL REVIEW(S)

TATISTICAL REVIEW

NDA:	21-303
SPONSOR:	Shire
DRUG:	Adderall XRTM CII (Extended Release Capsule — Mixed Salts of a Single-Entity Amphetamine Product)
INDICATION:	ADHD
MATERIAL SUBMITTED:	Statistical Review
DATE SUBMITTED:	10/03/2000
MEDICAL OFFICER:	Andy Mosholder, M.D.

I. BACKGROUND

The sponsor submitted an original NDA for a modified release mixed salt amphetamine product (SLI 381) to support a claim on Attention Deficit Hyperactivity Disorder (ADHD). Two studies were submitted for this NDA :

Study 381.301: "A Randomized, Double-Blind, Placebo-Controlled, Parallel Group Study of SLI381 in Children With Attention Defficit Hyperactivity Disorder", and

Study 381.201 : "A Randomized, Double-Blind, Placebo- and Active-Controlled, Crossover Study of SLI381 in Children with Attention Deficit Hyperactivity Disorder".

II. Study 381.301

This study was conducted from October 20 1999 to May 9, 2000. A total of 47 sites participated in the study.

II.1 Study Design

This was a multi-center, double-blind, randomized, parallel-group study of children aged 6 to 12 years old. This study included one-week single-blind placebo wash out period and three-weeks double-blind treatment period. Six weekly visits were scheduled: Visit 1 was used to screen subjects; Visit 2 was used to dispense single blind placebo; Visit 3 was the baseline visit to dispense double-blind treatment and Visits 3-5 were used to assess double-blind treatment. Patients were randomized to treatment groups in a 2:1 ratio (of treatment versus placebo, or in a 2:2:2:3 ratio of SLI381 10 mg: 20 mg: 30 mg: placebo). Daily morning dose of active drug or placebo was taken for three weeks. All patients started at 10 mg for the first week and patients who were assigned to 20 mg or 30 mg were titrated up to the full dose at week 2 and week 3, respectively. This is a forced

titration design, so dose modification based on efficacy evaluation or impressions was not allowed.

A block randomization schedule was implemented in this study. The block size is nine including 6 numbers for the SLI381 group and 3 numbers for placebo. The randomization schedule was not stratified by sites, but one complete randomization block package was sent to each site during the initial randomization.

II.2 Objectives

The primary objective of this study was to evaluate the efficacy and safety of SLI381 administered as a daily single morning dose. The secondary objective was to assess the diurnal efficacy of SLI381 based on the morning and afternoon evaluation.

II.3 Efficacy Endpoint

The primary efficacy endpoint was the average of the Conners 10-item Global Index Scale (CGIST, teacher's version), obtained during the last week of evaluation. The score for each item in the CGIST ranged from 0 to 3 ("not true at all" to "very much true"). Teacher rated the CGIST scores three times a week (Monday, Wednesday and Friday) in the morning and afternoon (10:00 a.m. and 2:00 p.m.) by contacting a call center.

Parents also completed the CGIS (CGISP) once per week on either Saturday or Sunday three times during the day (at approximately 10:00 a.m., 1:00 p.m. and 4:00 p.m.).

Secondary efficacy measure included the CGISP, Clinical Global Impressions Scale for severity (CGI-S) and change (CGI-C) and the parent's global assessment (PGA).

II.4 Number of Subjects and Analysis Plan

A total of 360 patients were planned to be enrolled (240 were allocated to SLI381 and 120 to placebo). This sample size was obtained to detect a 0.47 standardized treatment difference between SLI381 treatment versus placebo with 90% power, assuming a two-tailed 0.05 significance level. The standardized treatment difference of 0.47 can be converted to a raw score difference of 2.82 assuming a standard deviation of 6 points for the CGIST scores. Assuming 20% attrition rate, the placebo indicated that it is necessary to randomize approximately 450 subjects.

Intent-to-treat (ITT) population was used for all analyses. The ITT population was defined as all randomized patients who had at least one efficacy evaluation after randomization. The secondary population involved the per-protocol population and the

study completers. The per-protocol population included all randomized patients whose average drug compliance assessed by the pill count is at least 80% post randomization. Since the sponsor had randomized much more patients than that of originally planned (n=450), the agency requested that an additional analysis on the primary efficacy endpoint should be performed for the first 450 patients.

The protocol specified primary efficacy analysis was based on the one-way Analysis of Covariance Model (ANCOVA) using the CGIST scores average over a patient's last week measures as the dependent variable and the treatment (4 dose levels), the baseline score as the independent variables. Note that the two-way ANCOVA was used for the primary analysis in the report including site as an additional factor. Dunnett's test was used to control for comparison of each dose level versus placebo.

The average CGIST scores were calculated based on 6 values of a patient's last week assessments (one in the morning and one in the afternoon on each of three days : Monday, Wednesday and Friday). Similarly, the average baseline scores were computed based on the 6 assessments obtained from the placebo wash out period. If the third week efficacy score was missing, a LOCF was employed to impute the missing efficacy score. In the sponsor's original protocol (dated September 30, 1999), a standardized CGIST scores (a deviation of the raw scores from the site specific mean score divided by the site specific standard deviation) will be used for the ANCOVA model if a non-homogeneity of variance are noted. However, the sponsor's statistician indicated in the pre-NDA meeting (August 16, 2000) that this analysis will not be conducted.

As specified in the sponsor's analysis plan (dated June 20, 2000), based on the primary analysis result, the morning and afternoon CGIST total score were also analyzed separately to evaluate the prolonged action of the delayed release formulation. The same ANCOVA model as the primary endpoint analysis was used for the separated morning/afternoon CGIST total score analysis.

The sponsor planned to perform the primary efficacy analysis by subgroup including gender, age (6-9 vs. 10-12) and for patients with the same disease diagnosis. Based on the agency's request (due to the concern of randomizing more patients as planned), the sponsor also performed the primary efficacy analysis based on the first 450 patients (that was planned in the original protocol).

The sponsor also indicated that a secondary efficacy analysis on the parent's rated CGIS scores (CGISP) obtained during the last treatment week would be performed. The sponsor expected to demonstrate that the treatment would continue to be effective until the early evening.

To examine the treatment effect time course, the sponsor used the similar ANCOVA model as the primary analysis with the Dunnett's test for multiple means comparisons to analyze the mean CGIST total score for each treatment week.

For CGIST and CGISP scores, the missing items were evaluated for the morning and afternoon assessments and imputed with the mean score of the corresponding assessment if there are no more than two items with missing data. If there were more than two items with missing data, the total scores of the assessment were set to missing. In the statistical analysis plan, the sponsor indicated that patients would be excluded from the ANCOVA model if a patient had missing data on the baseline score. However, in the study report, the sponsor modified the statement as that "the missing baseline scores were imputed with the subjects' site-specific baseline mean".

II.5 Sponsor's Results

A total of 584 patients were randomized to the double-blind treatment phase. Seventy-five of these randomized patients did not complete the study. The most frequent reason for early discontinuation was withdrawn consent. Placebo had higher percent of patients discontinued, mostly due to withdrawn consent and lack of efficacy. Twenty-one patients who had missing post randomization CGIST score were not included in the ITT population. There are 120, 112, 128 and 203 ITT patients in 30 mg, 20 mg, 10 mg and placebo group, respectively. The patient disposition information was shown in Table I.A.1.

Table I.A.1 Patient Disposition

	Entire Study	Drop-out prior to Randomization	Randomized Treatment			
			30 mg	20 mg	10 mg	Placebo
Entered	649	65				
Randomized	584	N/A	124	121	129	210
Completed	509	N/A	112	105	119	173
Discontinued	140	65	12 (9.7%)	16 (13.2%)	10 (7.8%)	37 (17.6%)
By category						
AEs (death)	17	2(0)	5(0)	4(0)	0(0)	6(0)
Withdrawn consent	53	27	4	7	4	11
Protocol violation	13	7	2	1	2	1
Lost to follow-up	11	3	0	2	0	6
Lack of efficiency	13	N/A	0	1	2	10
Other	33	26	1	1	2	3
# of patients for assessment						
Efficacy (ITT)	563	N/A	120	112	128	203
Safety	649	65	124	121	129	210

Table I.A.2 shows the demographic and baseline information for the ITT population. The distribution of the gender, race, age, height, weight, diagnosis, disease duration, comorbid disease, baseline CGIST and CGISP scores appear to be comparable across treatment groups. Among 563 patients, 77.1% were male and 22.9% were female. 76.4% of these

patients were white. More than 91% of these patients had combined subtypes of ADHD diagnosis. Also, approximately 63% of the patients had prior treatment experience.

Table I.A.2 The Demographic and Baseline Information for the ITT Population

Characteristic	Treatment Group			
	SLI381 30 mg	SLI381 20 mg	SLI381 10 mg	Placebo
ITT (n=563)	120	112	128	203
Sex : Male Female	96(80.0%) 24(20.0%)	90(80.4%) 22(19.6%)	100(78.1%) 28(21.9%)	148(72.9%) 55(27.1%)
Race: Caucasian Black Hispanic Asian/PI Native American Other	84(70.0%) 20(16.7%) 11(9.2%) 1(0.8%) 0(0.0%) 4(3.3%)	92(82.1%) 9(8.0%) 10(8.9%) 0(0.0%) 0(0.0%) 1(0.9%)	98(76.6%) 11(8.6%) 12(9.4%) 1(0.8%) 1(0.8%) 5(3.9%)	156(76.8%) 27(13.3%) 15(6.9%) 2(1.0%) 0(0.0%) 4(2.0%)
Age (yr) mean	8.8 ± 1.77	8.4 ± 1.70	8.5 ± 1.58	8.6 ± 1.73
Hight (in) mean	53.1 ± 4.99	52.3 ± 4.52	52.3 ± 4.18	52.9 ± 4.38
Weight (lb) mean	72.6 ± 23.01	70.9 ± 24.09	73.2 ± 25.60	74.4 ± 25.84
Diagnosis: Combined Hyperactive Inattentive	112(93.3%) 6(5.0%) 2(1.7%)	104(92.9%) 6(5.4%) 2(1.8%)	117(91.4%) 8(6.3%) 3(2.3%)	190(93.6%) 8(3.9%) 5(2.5%)
Prior Rx: Amphetamine Methylphenidate Stimulant unspecified Other therapy No prior Rx Not listed	7(5.8%) 16(13.3%) 57(47.5%) 0(0.0%) 37(30.8%) 3(2.5%)	8(7.1%) 13(11.6%) 52(46.4%) 4(3.6%) 34(30.4%) 1(0.9%)	7(5.5%) 18(14.1%) 53(41.4%) 1(0.8%) 48(37.5%) 1(0.8%)	11(5.4%) 18(8.4%) 84(41.4%) 4(2.0%) 76(37.4%) 11(5.4%)
Comorbid conditions Yes No	37(30.8%) 83(69.2%)	31(27.7%) 81(72.3%)	41(32.0%) 87(68.0%)	61(30.0%) 142(70.0%)
CGIST-T Baseline score (N=539)	11.2 ± 6.9 (n=115)	12.1 ± 6.7 (n=108)	11.5 ± 6.3 (n=122)	10.6 ± 6.6 (n=194)
CGIS-P Baseline score (N=555)	12.9 ± 7.1 (n=117)	13.6 ± 7.5 (n=111)	12.4 ± 6.9 (n=127)	-13.8 ± 7.2 (n=200)

In the sponsor's primary analysis of CGIST total score using ANCOVA, the overall treatment effect and baseline score were highly significant (both had p-value<0.0001), but site effect was not significant (p-value=0.28, see Table I.A.3). Using the Dunnett's test to adjust for multiplicity, the CGIST total score showed significant improvement in all SLI381 dose groups in comparison to placebo (i.e. all SLI381 dose groups had significant lower mean CGIST total scores as compared with placebo; all p-values were <0.001).

All SLI381 dose groups also showed significant improvement over placebo at study endpoint based on a separate analysis of the morning or afternoon CGIST total score evaluation (i.e. based on a separate Dunnett test for the morning and afternoon measurements, all p-values were <0.001).

Table I.A.3 Efficacy Analysis of CGIS-Teacher at Study Endpoint

Efficacy parameter	Treatment Group			
	SLI381 30 mg	SLI381 20 mg	SLI381 10 mg	Placebo
Overall ITT (n=563)	120	112	128	203
Mean score #	4.6 ± 4.9 [-6.4,-3.9] **	6.1 ± 5.8 [-5.0,-2.4] **	6.2 ± 5.0 [-4.8,-2.3] **	9.8 ± 6.8
Mean diff from baseline	-6.4 ± 6.2	-6.0 ± 6.1	-5.3 ± 5.3	-0.9 ± 5.0
Morning ITT (n=560)	120	112	128	200
Mean score #	4.9 ± 5.4 [-5.8,-3.1] **	6.2 ± 5.9 [-4.6,-1.8] **	6.2 ± 5.2 [-4.5,-1.9] **	9.3 ± 6.8
Mean diff from baseline	-5.8 ± 6.6	-5.4 ± 6.4	-5.0 ± 5.9	-0.7 ± 5.2
Afternoon ITT (n=560)	120	112	128	200
Mean score #	4.3 ± 4.9 [-7.1,-4.5] **	5.9 ± 5.9 [-5.7,-2.9] **	6.4 ± 5.2 [-5.2,-2.5] **	10.2 ± 7.3
Mean diff from baseline	-7.2 ± 6.7	-6.8 ± 6.7	-5.4 ± 5.4	-1.2 ± 5.4

Note : # Average of total morning and afternoon scores during the last treatment week

** p-value <0.001 compared to placebo by Dunnett test following 2-way ANCOVA

[] 95% confidence interval with reference to placebo calculated by Dunnett's test following 2-way ANCOVA

The results of the parent evaluation of CGIS (CGISP) showed similar significant efficacy (Table I.A.4). The overall, morning and afternoon mean CGISP total scores for all

SLI381 dose groups were significantly lower than the scores for the placebo group based on the Dunnett's tests.

Table I.A.4 Efficacy Analysis of CGIS-Parent at Study Endpoint

Efficacy parameter	Treatment Group			
	SLI381 30 mg	SLI381 20 mg	SLI381 10 mg	Placebo
Overall ITT (n=558)	120	110	128	200
Mean score #	6.8 ± 6.0 [-6.7,-3.5] **	8.5 ± 6.8 [-5.1,-1.7] **	8.7 ± 6.8 [-4.8,-1.6]**	11.9 ± 7.4
Mean diff from baseline	-6.1 ± 7.7	-5.0 ± 7.0	-3.7 ± 7.6	-2.0 ± 6.3
Morning ITT (n=552)	119	109	128	196
Mean score	6.8 ± 6.1 [-6.4,-3.0] **	8.4 ± 7.3 [-4.9,-1.3] **	8.9 ± 7.3 [-4.2,-0.9] *	11.5 ± 7.6
Mean diff from baseline	-5.7 ± 8.5	-4.7 ± 7.6	-3.5 ± 7.7	-1.7 ± 7.0
Afternoon ITT (n=557)	120	110	128	199
Mean score	6.2 ± 6.3 [-7.4,-3.9] **	7.8 ± 7.2 [-5.7,-2.0] **	8.3 ± 7.1 [-5.1,-1.6] **	11.7 ± 7.8
Mean diff from baseline	-7.0 ± 8.0	-5.1 ± 7.9	-4.3 ± 8.4	-2.0 ± 7.2
Evening ITT (n=554)	119	110	126	199
Mean score	7.6 ± 6.8 [-7.3,-3.5] **	9.6 ± 8.3 [-5.9,-2.1] **	8.9 ± 7.5 [-5.3,-1.4] **	12.9 ± 8.0
Mean diff from baseline	-5.8 ± 8.1	-4.7 ± 8.9	-3.7 ± 9.3	-1.4 ± 7.7

Note : # Average of total morning and afternoon scores during the last treatment week

** p-value <0.001 compared to placebo by Dunnett's test following a 2-way ANCOVA

* p-value <0.01 compared to placebo by Dunnett's test following a 2-way ANCOVA

[] 95% confidence interval with reference to placebo calculated by Dunnett's test following 2-way ANCOVA

The sponsor evaluated the CGIST total score by various subgroups : gender, prior stimulant therapy and response of the first 450 patients randomized. In these analyses, the sponsor found out that the treatment effect was different between boys and girls (Table I.A.5), specifically, the CGIST total scores were reduced more in boys than those in girls. The sponsor indicated that the different treatment between gender may be attributed to the larger placebo effect occurred in girls (the reduction in CGIST total scores at endpoint from baseline for placebo group was significant in girls : p-

value=0.0004, but not in boys : p-value=0.3354). The placebo effect may contribute to the smaller SLI381 benefit (as compared to the placebo group) in girls than that in boys. The sponsor also noted that the number of girls was small compared with the number of boys and this may affect the statistical testing result. The sponsor found out that the treatment effect in stimulant experienced patients and treatment-naïve patients were very similar. The analysis result based on the first 450 randomized patients was also very similar to that based on all ITT population.

Table I.A.5 Efficacy Analysis of CGIS-Teacher at Study Endpoint by subgroups

Efficacy parameter	Treatment Group			
	SLI381 30 mg	SLI381 20 mg	SLI381 10 mg	Placebo
Boys ITT (n=434)	96	90	100	148
Mean score #	4.7 ± 5.1	6.0 ± 5.9	6.6 ± 5.0	11.9 ± 7.4
Mean diff from baseline	-6.8 ± 6.2	-6.6 ± 5.9	-5.2 ± 5.3	-0.4 ± 5.1
Girls ITT (n=129)	24	22	28	55
Mean score #	4.3 ± 3.8	6.3 ± 5.4	4.7 ± 4.5	7.0 ± 5.3
Mean diff from baseline	-5.1 ± 5.7	-3.9 ± 6.7	-5.2 ± 5.3	-2.2 ± 4.4
Stimulant experienced ITT (n=343)	80	73	78	112
Mean score #	5.0 ± 5.0	6.2 ± 6.0	6.9 ± 5.3	10.9 ± 6.9
Mean diff from baseline	-6.9 ± 6.4	-6.1 ± 6.5	-4.6 ± 5.4	-0.8 ± 4.5
Naive patient ITT (n=193)	37	33	47	76
Mean score #	4.1 ± 4.5	5.8 ± 5.4	5.2 ± 4.3	8.6 ± 6.5
Mean diff from baseline	-5.6 ± 5.9	-5.9 ± 5.7	-6.1 ± 5.2	-0.4 ± 5.5

Although the sponsor did not specify the by-week analysis of CGIST as their primary objective in the SAP or protocol, the sponsor wants to show the first-week treatment effect in their labeling anyway based on the by-week analysis. They performed the ANCOVA analyses of the CGIST total scores (including treatment, center, baseline CGIST scores in the model), based on the overall, morning and afternoon data, by week, using the observed case data. Then they used the Dunnett's test to adjust for the multiple comparison between each dose level and placebo at each treatment week. They did not consider forced titration design for the treatment comparison. Based on this design, the different sets of dose level and placebo were compared at each treatment week. For example, at first week, only 10 mg dose level was compared with placebo; at week 2, 10 mg and 20 mg dose level were compared with placebo; and at week 3, 10 mg, 20 mg and 30 mg were compared with placebo. In stead of doing so, they simply compared each dose level (as randomized) with placebo at each time point. Based their analysis, they obtained significant result ($p < 0.001$) for almost all the pair-wise comparison (Table A.I.6).

The sponsor also performed the similar analysis for the first 450 patients for the overall data. The results were very consistent with the overall results presented here.

Table I.A.6 Efficacy Analysis of CGIS-Teacher by Treatment Week

Efficacy parameter	Treatment Group			
	SLI381 30 mg	SLI381 20 mg	SLI381 10 mg	Placebo
Overall ITT (n=563)	120	112	128	203
Week 1 OC : N (n=539)	115	108	122	190
Diff between Mean score #	-3.2 [-4.3,-2.0] **	-2.8 [-4.0,-1.7] **	-2.7 [-3.9,-1.5]**	
Week 2 OC : N (n=508)	113	103	114	178
Diff between Mean score #	-4.2 [-5.4,-2.9] **	-3.5 [-4.8,-2.2] **	-2.7 [-3.9,-1.4] **	
Week 3 OC : N (n=469)	101	98	115	155
Diff between Mean score #	-4.9 [-6.2,-3.5] **	-3.5 [-4.8,-2.1] **	-3.2 [-4.5,-1.9] **	

Table I.A.6 Efficacy Analysis of CGIS-Teacher by Treatment Week (continued)

Efficacy parameter	Treatment Group			
	SLI381 30 mg	SLI381 20 mg	SLI381 10 mg	Placebo
Morning				
ITT (n=560)	120	112	128	200
Week 1 OC : N (n=534)	116	107	124	187
Diff between Mean scores #	-4.4 [-5.8,-3.1] **	-3.2 [-4.6,-1.8] **	-3.2 [-4.5,-1.9] **	
Week 2 OC : N (n=506)	113	103	113	177
Diff between Mean scores #	-3.1 [-4.5,-1.8]**	-2.8 [-4.2,-1.3]**	-1.9 [-3.3,-0.5]*	
Week 3 OC : N (n=466)	101	97	115	153
Diff between Mean scores #	-4.2 [-5.6,-2.7]**	-2.9 [-4.4,-1.4]**	-2.8 [-4.2,-1.4]**	
Afternoon				
ITT (n=560)	120	112	128	200
Week 1 OC : N (n=534)	116	106	123	189
Diff between Mean scores #	-4.0 [-5.3,-2.8] **	-3.3 [-4.5,-2.0] **	-3.2 [-4.5,-1.9] **	
Week 2 OC : N (n=502)	112	103	111	176
Diff between Mean scores #	-5.1 [-6.4,-3.7]**	-4.1 [-5.5,-2.7]**	-3.4 [-4.8,-2.1]*	
Week 3 OC : N (n=465)	101	98	113	153
Diff between Mean scores #	-5.7 [-7.2,-4.2]**	-4.1 [-5.6,-2.6]**	-3.6 [-5.0,-2.1]**	

Note : # Difference of the CGIST total score between each dose level and placebo

** p-value <0.001 compared to placebo by Dunnett test following a 2-way ANCOVA

* p-value <0.01 compared to placebo by Dunnett test following a 2-way ANCOVA

[] 95% confidence interval with reference to placebo calculated by Dunnett's test following 2-way ANCOVA

• There was no imputation for missing baseline CGIST total scores.

II.6 Reviewer's Evaluation and Comments

This reviewer performed the sponsor's primary analysis according to the analysis plan (two-way ANCOVA model using average CGIST total scores at study endpoint and with LOCF) and obtained the significant overall treatment effect ($p=0.0001$). After Dunnett's adjustment, all pair-wise comparisons were also significant. The results show the significant treatment effect in favor of the three SLI381 dose groups by larger reduction in the CGIST total scores as compared with those from placebo. A slight inconsistency of the sponsor derived data and the reviewer created data was found, but since the difference was small, it did change the study results. For example, among a total of 563 observations for the independent variable (averaged CGIST total scores), only 4 inconsistent data points were found; among a total of 563 observations for the baseline average CGIST total score, only 7 inconsistent data points were found.

The reviewer also conducted the sponsor's primary analysis by the morning and afternoon assessments and the results were also very similar to the sponsor's finding. Again, a slight discrepancy in the sponsor's derived data and the reviewer's created data was found, but it did not affect the study result.

In the separated analysis (by morning and afternoon assessments), the sponsor tried to show that the delayed release formulation had a prolong effect from morning through the afternoon. But the sponsor's intention to incorporate this analysis as their primary analysis result was not clear. In their statistical analysis plan (SAP, final version dated June 20, 2000), this evaluation was listed as the secondary endpoint (see "Study Objective" section). However, in the SAP "Analysis of Primary Efficacy Endpoints" section, it indicated that, based on the results of the ITT analyses, the ANCOVA model would be applied to patients' average of CGIST total scores obtained at morning and afternoon assessments. Based on this reviewer's reading, this is the only statement that indicated their intention to include the morning-afternoon assessments in the primary analysis.

In addition, although the sponsor proposed the Dunnett's test to adjust for the comparisons between each dose level versus placebo for the primary endpoint (the average CGIST total scores at the last week measures), the decision rule with regard to the additional endpoints (separate morning and afternoon measures) and pair-wise comparison within each endpoint were not mentioned. However, this reviewer applied the conditional testing approach starting with testing for the overall effect (morning and afternoon). If after Dunnett's adjustment, all comparisons between dose level and placebo were significant for the overall effect, then the similar testings were performed for the morning evaluation; if all comparisons were significant, the procedure would go further to test for the afternoon evaluation. Based on this procedure, this reviewer found that the sponsor's results demonstrated significant treatment effect in favoring of the three SLI381 dose groups based on a separated Dunnett's test for both morning and afternoon evaluations.

This reviewer performed additional analyses based on the primary analysis (with treatment, center and baseline score as independent variables) with center by treatment interaction, or with treatment by baseline score interaction or with gender, treatment by gender interaction included. The treatment effect remained significant in these analyses. However, two significant treatment by factor interactions were found : treatment by gender interaction and treatment by baseline score interaction. The treatment by gender interaction was discussed in the sponsor's report that the boys had more reduction in CGIST total scores than the girls. But the sponsor attributed the finding to the sample size that it was too small for girls to draw definitive conclusion. This reviewer dichotomized the baseline CGIST total scores based on the median baseline score and found that patients who had baseline scores higher than the median would have more improvement than patients who had baseline scores lower than the median. However, all SLI381 dose groups seem to have statistically significant improvement over placebo group in both subgroups (patients above or below the median baseline score) (see Table I.B.1).

Table I.B.1 Efficacy Analysis of CGIS-Teacher at Study Endpoint by Baseline CGIS-T Score Status (\leq median score; $>$ median score)

Efficacy parameter	Treatment Group			
	SLI381 30 mg	SLI381 20 mg	SLI381 10 mg	Placebo
\leq median score ITT (n=286)	64	50	61	111
Mean score #	3.70 \pm 3.45	3.66 \pm 4.05	2.95 \pm 3.38	6.17 \pm 4.61
Difference between means	-3.22	-2.51	-2.47	
95% C.I.	[-4.74, -1.70]*	[-4.16, -0.86]*	[-4.01, -0.92]*	
$>$ median score ITT (n=277)	56	62	67	92
Mean score #	8.51 \pm 5.09	8.02 \pm 6.21	6.53 \pm 5.57	14.07 \pm 6.39
Difference between means	-7.54	-6.05	-5.55	
95% C.I.	[-9.94, -5.14]*	[-8.38, -3.73]*	[-7.83, -3.28]*	

Note : # Average of total morning and afternoon scores during the last treatment week

* p-value < 0.05 compared to placebo by Dunnett's test

[] 95% confidence interval with reference to placebo calculated by Dunnett's test

DSMB meeting had been conducted. Analysis was performed by blinded group assignment. Since sample sizes were not equal among active treatment groups and placebo, it is not clear how the analysis can be performed by blinded group.

With respect to the time course of treatment effect overtime (i.e. by study week), this reviewer performed the similar 2-way ANCOVA (except using LOCF at each time point)

based on sequential testing approach for the treatment effect overtime, starting from treatment week 3 and working backward. The sequential testing approach is described as follows :

1. At week 3, given the significant overall treatment effect (10 mg, 20 mg, 30 vs. placebo), a Dunnett's test was performed to adjust for the pair-wise comparisons. Only if all the results were significant at 0.05 level for the overall, morning or afternoon effect, the procedure would go to the next step; otherwise, the procedure would stop.
2. At week 2, given the significant overall treatment effect (10 mg, 20 mg vs. placebo), a Dunnett's test was performed to adjust for the pair-wise comparisons. Only if all the results were significant at 0.05 level for the overall, morning or afternoon effect, the procedure would go to the next step; otherwise, the procedure would stop.
3. Similarly, at week 1, given the significant overall treatment effect (10 mg vs. placebo), a Dunnett's test was performed to adjust for the pair-wise comparisons. Only if all the results were significant at 0.05 for the overall, morning or afternoon effect, the procedure would declared that the overall, morning or afternoon treatment effect at week 1 were significant (Note, only 10 mg was compared with placebo at this week).

Table I.B.2.1 Efficacy Analysis of CGIS-Teacher for Week 3

Efficacy parameter	Treatment Group			
	SLI381 30 mg	SLI381 20 mg	SLI381 10 mg	Placebo
WEEK 3				
Overall				
LOCF:N (n=563)	120	112	128	203
Lsmean ♦	4.4 **	5.4 **	5.8 **	9.8
Diff between Mean scores #	-5.1 [-6.4,-3.9] *	-3.7 [-5.0,-2.4] *	-3.5 [-4.8,-2.3] *	
Morning				
LOCF:N (n=560)	120	112	128	200
Lsmean ♦	4.7 **	5.6 **	5.7 **	9.4
Diff between Mean scores #	-4.4 [-5.8,-3.1] *	-3.2 [-4.5,-1.8] *	-3.2 [-4.5,-1.9] *	
Afternoon				
LOCF:N (n=558)	119	111	127	201
Lsmean ♦	4.1 **	5.1 **	6.1 **	10.1
Diff between Mean scores #	-5.9 [-7.3,-4.5] *	-4.3 [-5.7,-2.9] *	-3.8 [-5.2,-2.5] *	

Note : ♦ Least squared means (Lsmean) from 2-way ANCOVA.

** p-value=0.0001 which p-value was from the comparison of Lsmean between each dose level and placebo (without multiplicity adjustment)

Difference of the mean CGIST total score between each dose level and placebo

* p-value <0.05 compared each dose level to placebo by Dunnett's test

[] 95% confidence interval with reference to placebo calculated by Dunnett's test

• The center specific mean CGIST total score was used for the missing baseline CGIST total score.

• Treatment effect in the ANCOVA models was significant (p=0.0001).

Table I.B.2.2 Efficacy Analysis of CGIS-Teacher For Week 2

Efficacy parameter	Treatment Group		
	SLI381 20 mg	SLI381 10 mg	Placebo
WEEK 2			
Overall			
LOCF:N(n=562)	232	127	203
Lsmean ♦	5.7 **	6.4 **	10.3
Diff between Mean scores #	-4.2 [-5.2,-3.3] *	-3.4 [-4.5,-2.2] *	
Morning			
LOCF:N(n=558)	232	126	200
Lsmean ♦	5.7 **	6.2 **	9.7
Diff between Mean scores #	-3.6 [-4.6,-2.6] *	-2.8 [-4.0,-1.6] *	
Afternoon			
LOCF:N(n=555)	230	125	200
Lsmean ♦	5.6 **	6.4 **	10.9
Diff between Mean scores #	-4.9 [-5.9,-3.9] *	-4.1 [-5.3,-2.9] *	

Note : Same as Table I.B.2.1

- Treatment effect in the ANCOVA models was significant ($p=0.0001$).

Table I.B.2.3 Efficacy Analysis of CGIS-Teacher For Week 1

Efficacy parameter	Treatment Group	
	SLI381 10 mg	Placebo
WEEK 1		
Overall		
LOCF:N(n=539)	349	190
Lsmean ♦	6.4 **	9.8
Diff between Mean scores #	-2.9 [-3.7,-2.2] *	
Morning		
LOCF:N(n=534)	347	187
Lsmean ♦	6.2 **	9.3
Diff between Mean scores #	-2.5 [-3.3,-1.7] *	
Afternoon		
LOCF:N(n=534)	345	189
Lsmean ♦	6.5 **	10.4
Diff between Mean scores #	-3.5 [-4.3,-2.7] *	

Note : Same as Table I.B.2.1

- Treatment effect in the ANCOVA was significant ($p=0.0001$).

APPEARS THIS WAY
ON ORIGINAL

These analyses were shown at Tables I.B.2.1, I.B.2.2 and I.B.2.3 for steps 1, 2 and 3, respectively. Note that the p-value based on the comparison of least square means was 0.0001 for all pair-wise tests which did not have multiplicity adjustment. Using the most conservative approach, the p-value could be as large as 0.0003 with three pair-wise comparisons.

Based on the Dunnett's test, all the pair-wise comparisons of the overall, morning/afternoon treatment effect were significant at week 3. So the tests were further performed at week 2. Again, since all comparisons were significant. The step 3 procedure for week 1 was performed. The result shows that the 10 mg group was significantly better than placebo at week 1.

APPEARS THIS WAY
ON ORIGINAL

APPEARS THIS WAY
ON ORIGINAL

III. Study 381.201

This study was conducted from September 21, 1999 to December 4, 1999. A total of 4 sites participated in the study. The final statistical analysis plan (SAP) was dated February 14, 2000. The final study report was dated July 31, 2000.

III.1 Study Design

This was a randomized, multicenter, double-blind, placebo and active-controlled, crossover, multiple-dose trial of three doses of SLI381 versus Adderall versus placebo in subjects who had been diagnosed with ADHD according to the DSM-IV criteria.

The study included a screening period, a single-dose practice classroom session (with pharmacokinetic measures), and six weeks of double-blind medication session (with pharmacokinetic measures performed again at the end of the sixth week). Each week represents a treatment period. The treatment periods consisted of SLI381 10 mg, 20 mg, 30 mg, Adderall 10 mg and placebo, with an extra double-blind week to adjust for any missing treatment period. During the treatment period, medication was administered once a day in the morning as an oral dose. Patients who successfully completed the open-label practice session were randomized to the double-blind treatment sequence.

The practice session was scheduled for the Saturday following the washout period. During the practice session, patients were given 20 mg of SLI 381 in the morning and the SKAMP (Swanson, Kotkin, Agler, M-Flynn, and Pelham Rating Scale) and PERMP (Permanent Product Measure of Performance) measures were completed at 0, 1.5, 3.0, 4.5, 6.0, 7.5, 9.0, 10.5, and 12.0 hours post-dose. The timing of these efficacy measures throughout the study closely matched the timing of blood draws for PK analysis. Raters were instructed to consistently observe the same children from week to week.

At the end of the practice session, subjects who tolerated the 20 mg dose of SLI381 were randomized to double-blind treatment sequence which set the order of the treatment to be taken for the next five weeks. Saturday classes continued for those five weeks. The sponsor indicated in the protocol that, for each week, subjects will be randomized to a different treatment sequence each Sunday until each treatment sequence has been completed by each subject. A Latin Square design was used to determine randomization of subjects with approximately one-fifth of subjects being randomized to each of the five sequences. On each Saturday, objective and subjective measures were evaluated. The measures included two factors of classroom behavior (attention and deportment) using SKAMP, a patient completed a series of math problems to obtain an objective measure of performance on problems attempted (PERMP Attempted) and on problems correct (PERMP Correct). Both SKAMP and PERMP measures were completed during each 1.5 hour classroom cycles (similar to the measures in the practice session).

At the end of the fifth week, a separate randomization procedure was used to determine the treatment for the make-up week to allow for random duplication of a single treatment condition for subjects who did not miss any treatment conditions. Otherwise, patient would take the medication that he/she missed during the double-blind period, at the make-up week. If a subject missed more than one medication, he/she would receive the first medication that was missed at this make-up week. However, the data in this period was not used in the analysis.

III.2 Objectives

The primary objective of the study was to assess the efficacy and safety of SLI381 10, 20 and 30 mg compared to placebo. The secondary objectives were to assess the morning and afternoon therapeutic responses to SLI381 and to examine the pharmacokinetic profile of SLI381.

III.3 Study Endpoint

In the protocol, the primary efficacy variable was the SKAMP teacher rating scale (attention, deportment) evaluated at each session (eight equally spaced intervals based on time since dose) and period. The primary comparison was between SLI381 groups and placebo. The SKAMP score included 13 items ranging from 0 (normal) to 6 (maximum impairment) and would be evaluated by independent observers. It was not clear whether the SKAMP (attention, deportment) total score or average score per item would be analyzed in the protocol or SAP. But in the report, the average SKAMP score per item was used in the primary efficacy analysis.

In the SAP and study report, SKAMP (attention, deportment) and PERMP (attempted, correct) were co-listed as the primary efficacy variables. **Note that both scales had two sub-scales. The sub-scales were analyzed separately.**

III.4 Number of Subjects and Analysis Plan

Approximately 45 subjects (aged 6-12 years) were randomized to achieve at least 30 completed subjects. Each center was requested to randomize up to 15 subjects to obtain a maximum of 60 subjects randomized and to achieve at least 36 completed.

The intent-to-treat population (defined as all randomized patients who had at least one efficacy data point in the first five weeks of double-blind treatment) was used for the primary efficacy analysis. However, this definition for ITT population that was usually used for parallel trial does not seem to be appropriate for the cross-over design. The per-protocol population (PP, defined as all randomized patients who had completed the first five weeks of double-blind treatment) was used for the secondary efficacy analysis.

In the original protocol (9/1/99), a repeated measures ANOVA model with session and treatment as independent variables, possibly, adjusting for baseline (practice session)

covariate was proposed for the primary analysis. The sponsor also proposed to replace the missing data by using the average values from adjacent time points to allow for repeated measures ANOVA on all subjects who completed the five medication conditions. In addition, in the original protocol, a closed testing procedure was proposed. The procedure starts with testing the SLI381 **combined dosage groups** versus placebo, then each of the dose level (from highest to lowest dose level and followed by active control adderall) versus placebo if the combined dose level was significantly better than placebo. There was no correction for multiple comparison. A 0.05 significant level was required for each comparison in order to declare success and proceed to the next level.

In the primary analysis proposed in the original protocol, there was no clear objectives about what time point or what combination of SKAMP scores (e.g. mean score over-time) will be compared. But in the "Secondary Efficacy Variables and Analysis" section (in page 23 of the protocol), it indicated that **the time effects by dose** will be analyzed. In this analysis, **the duration of action of SLI 381 within each test session**, time course effects for each medication condition was graphically depicted and analyzed for both the SKAMP (attention, deportment) and PERMP. Time of peak effects for each treatment group was obtained based on the similar repeated measures ANOVA models.

In the SAP (dated 2/14/00), a mixed effect model was used for the primary analysis (Note : this was different from the protocol specified repeated measured ANOVA model). In this model, the sponsor stated that subject nested within site was treated as the random effect and treatment, period, session and the treatment-by-session interaction were treated as the fixed effects. Contrasts were constructed to compare each of the SLI381 dose level with placebo and with adderall **within each session**. A similar closed testing procedure described earlier was applied here to account for multiple comparison. Note: the closed testing procedure was only applied to adjust for the multiple dose levels, not with regard to time points.

In addition to the previous stated analysis, the sponsor also indicated that a simple analysis for treatment difference was done for each session. The analysis was a single-factor ANOVA model without adjusting for multiple testing.

In the study report (dated 7/31/00) "Analysis of Efficacy Endpoints" section, the sponsor further indicated that the overall treatment effect (**averaged across the scores observed on the classroom day**) was tested first for significance. Based on the results of the mixed effect ANOVA, planned pair-wise comparisons of each of the active treatment versus placebo were performed **within each classroom session**, using linear contrasts. The sponsor tried to use the pattern or profile of the significant differences along the time domain as shown by these planned comparisons to indicate the onset and duration of drug effect during a school day which lasted for 12 hours.

Overall, this reviewer's impression was that the time course evaluation described in the SAP and study report as the primary efficacy analysis was originally listed as the secondary efficacy variables and analysis in the protocol.

III.5 Sponsor's Result

Due to the concern about the study design, lack of clear specification of the primary endpoint, validity of the primary efficacy analysis (see reviewer's evaluation section), only a brief review of the study results was provided.

There were 51 patients enrolled. All of these patients completed the practice visit and were randomized into the double-blind treatment : 92% patients completed the first 5 weeks of double-blind treatment; and 86% completed the final (makeup) week (week 6) treatment. Table II.A.1 summarized patient disposition :

Table II.A.1 Patient Disposition

	No. of subjects
Enrolled	51
Randomized	51
Completed #	44
Discontinued	7
AME	2
Withdrew consent	1
Lost to follow-up	2
Other _a	2
Analysis population	
Efficacy	
ITT _b	49
PP _c	47
Safety	51

APPEARS THIS WAY
ON ORIGINAL

Note : #: This only counted patients who completed all periods, not completed all time points within a period.

a : Subject 2-17 could not tolerate study; subject 3-04 experienced menarche.

b : Subjects 1-08 and 3-04 discontinued prior to any efficacy evaluations during the double-blind treatment.

c : Subjects 2-17 and 4-08 discontinued prior to completing first 4 weeks of double-blind treatment.

The summary of demographic and baseline characteristics were presented in Table II.A.2 in which all the treatment sequences were combined. From this table, it shows that more boys (86%) were randomized than girls (14%). The three major ethnic groups were white (49%), Hispanic (23.5%) and Black (15.7%). Mean age was 9.5 years old. More older kids (9-12 years old : 64.7%) were randomized than younger kids (6-8 years old : 35.3%). Most diagnosis was combined ADHD (98%). The majority of patients had prior treatment (92.1%).

Table II.A.2 Summary of Demographic and Baseline Characteristics

Characteristic	All patients (n=51) n(%)
Gender Male Female	44(86.3) 7 (13.7)
Race White Black Hispanic Asian/pacific islander Other	25(49.0) 8(15.7) 12(23.5) 3(5.9) 3(5.9)
Age (yr) Mean SD Min-Max	9.5 1.9 6.0-12.0
Age distribution 6-8 yr 9-12 yr	18(35.3) 33(64.7)
Weight (lb) Mean SD Min-Max	83.5 28.9 48.0-161.0
Height (in) Mean SD Min-Max *	54.6 4.9 45.5-65.2
Diagnosis of ADHD Hyperactive Combined	1(2.0%) 50(98.0%)
Duration of Treatment (Yr) Mean SD Min-Max	1.7 1.7 -----
Previous treatment Amphetamine only Methylphenidate only None	17(33.3%) 30(58.8) 4(7.8)

APPEARS THIS WAY
ON ORIGINAL

APPEARS THIS WAY
ON ORIGINAL

For efficacy results evaluation, only the SKAMP (attention and deportment) scores will be evaluated as the primary efficacy endpoints (i.e. protocol specified).

III.5.1 SKAMP Attention

Based on the sponsor's mixed effect model, the treatment, period and session effect were all highly significant ($p < 0.0001$). The treatment-by-session interaction was also significant ($p < 0.0001$).

The sponsor then performed pair-wise comparisons to compare SLI381 30 mg, 20 mg, 10 mg and Adderall 10 mg versus placebo by average over the 8-session scores. The average score over the 8 sessions were 1.51, 1.33, 1.30, 1.18 and 1.02 for placebo, Adderall 10 mg, SLI381 10 mg, 20 mg and 30 mg, respectively. The differences in the average score

were highly significant for each SLI381 dose levels and Adderall 10 mg versus placebo ($p < 0.0001$ for all SLI381 dose levels and Adderall 10 mg, except SLI381 10 mg, for which p is < 0.001).

The sponsor went further to perform pair-wise mean comparisons over the time course (at each session). The duration of effect was determined based on the time points that showed significant treatment difference in favor of treatment ($p < 0.05$) (Table II.A.3). The sponsor's summary of the time points with significant treatment effect were :

- SLI381 30 mg : All time points from 1.5 to 12 hours;
- SLI381 20 mg : Starting at 4.5 hours to 12 hours post dose; and marginal significance was seen at 1.5 hours post dose ($p = 0.0513$);
- SLI381 10 mg : at 4.5, 6, 7.5 and 10.5 hours post dose; and marginal significance was seen at 12.0 hours post dose ($p = 0.0626$);
- Adderall 10 mg : at 1.5, 4.5, 6.0 and 7.5 hours post dose.

A graphical presentation of the mean SKAMP attention score by treatment and session was shown in figure II.A.1.

Figure II.A.1 Mean SKAMP Attention Score by Treatment and Classroom Session (ITT)

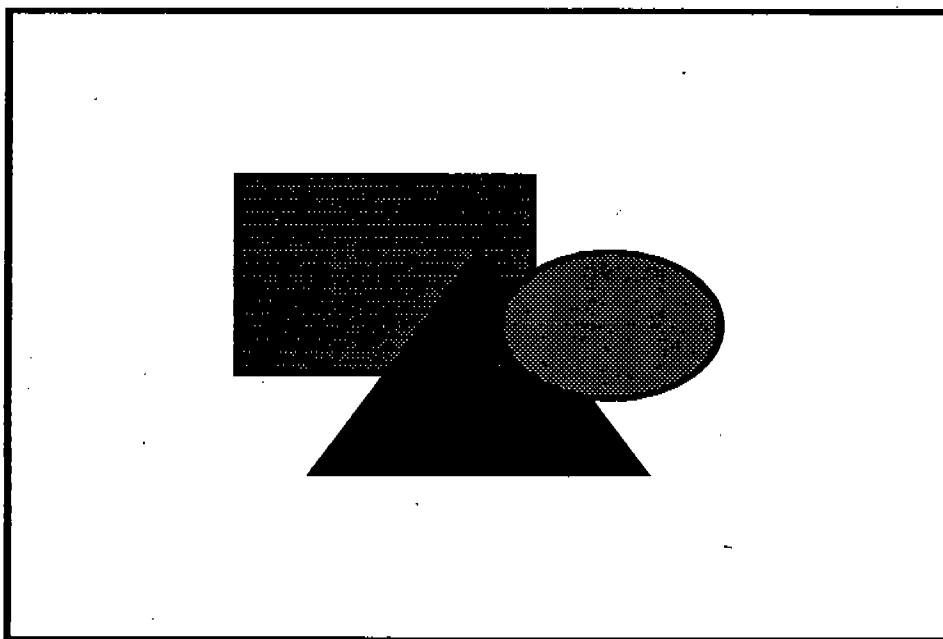


Table II.A.3 Mean Scores (SD) of SKAMP Rating Scale : Attention

Time (hr) Post dose	Parameter P-value #	Treatment				
		Placebo	Adderall 10 mg	SLI381 10 mg	SLI381 20 mg	SLI381 30 mg
0.0	N	44	44	41	44	41
	Mean	1.18	1.59**	1.55*	1.27	1.38
	SD	0.98	1.21	1.05	1.15	1.08
	P=0.34					
1.5	N	44	43	42	45	42
	Mean	1.31	0.88**	1.27	1.16	0.98**
	SD	0.83	0.83	0.93	1.09	0.99
	P=0.16					
4.5	N	44	42	42	44	42
	Mean	1.40	0.92***	1.13*	1.07**	0.90***
	SD	0.95	0.92	0.98	1.09	0.82
	P=0.11					
6.0	N	44	42	42	45	42
	Mean	1.74	1.26***	1.26***	1.14***	0.74***
	SD	1.01	1.18	1.03	0.99	0.74
	P=0.0003					
7.5	N	44	43	42	45	42
	Mean	1.73	1.22***	1.21***	1.13***	0.74***
	SD	1.01	0.87	0.80	1.05	0.81
	P=0.0001					
9.0	N	44	43	41	45	42
	Mean	1.51	1.55	1.40	1.26**	1.05***
	SD	0.97	1.28	1.02	1.15	1.16
	P=0.22					
10.5	N	42	42	40	45	41
	Mean	1.74	1.60	1.40**	1.27***	1.23***
	SD	0.87	1.26	1.02	0.88	1.28
	P=0.12					
12.0	N	42	42	40	44	41
	Mean	1.44	1.59	1.23	1.18**	1.15**
	SD	0.93	1.21	0.96	1.01	1.21
	P=0.24					

Note : # : p-value was based on ANOVA analysis by session for treatment difference (treatment was the only factor)

* p<0.05 : compared with placebo using mixed-effect ANOVA

** p<0.01 : compared with placebo using mixed-effect ANOVA

*** p<0.001 : compared with placebo using mixed-effect ANOVA

III.5.2 SKAMP Department

Similar to the SKAMP attention score results, the treatment, session, period effects and treatment-by-session interaction were all highly significant ($p < 0.0001$) based on the mixed-effect ANOVA model.

The pair-wise comparison of average score over the 8 sessions for each SLI381 dose level and Adderall 10 mg versus placebo were also highly significant ($p < 0.0001$).

Using the similar analysis method (as described earlier for SKAMP attention score), the sponsor performed pair-wise mean comparisons over the time course (Table II.A.4). The significant treatment effects over time course were summarized as follows (again, the significance was declared when $p < 0.05$):

- SLI381 30 mg : at all time points from 1.5 to 12 hours post dose;
- SLI381 20 mg : Starting at 1.5 hours to 10.5 hours post dose; and marginal significance was seen at 12 hours post dose ($p = 0.0531$);
- SLI381 10 mg : at 4.5, 6, 7.5 and 9 hours post dose; and marginal significance was found at 1.5 hours ($p = 0.0725$) and 10.5 hours post dose ($p = 0.0724$);
- Adderall 10 mg : at all time points from 1.5 to 10.5 hours post dose.

A graphical presentation of the mean SKAMP attention score by treatment and session was shown in figure II.A.2.

Figure II.A.2 Mean SKAMP Department Score by Treatment and Classroom Session (ITT)

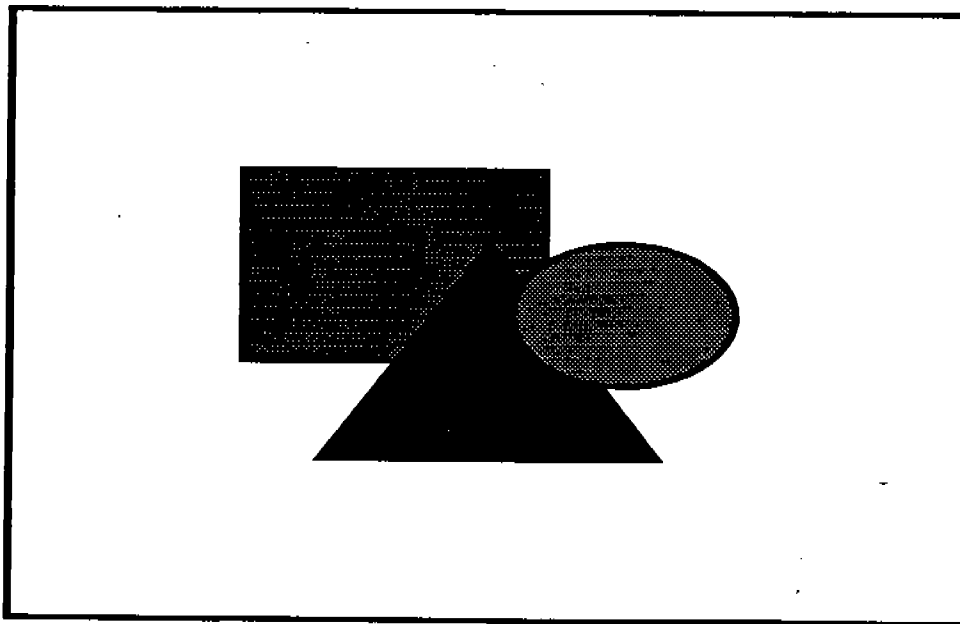


Table II.A.4 Mean Scores (SD) of SKAMP Rating Scale : Deportment

Time (hr) Post dose	Parameter P-value #	Treatment				
		Placebo	Adderall 10 mg	SL1381 10 mg	SL1381 20 mg	SL1381 30 mg
0.0	N	44	44	41	44	41
	Mean	1.88	2.43**	2.28	2.26	1.96
	SD	1.39	1.76	1.50	1.37	1.51
	P=0.40					
1.5	N	44	43	42	45	42
	Mean	2.22	1.08***	1.91	1.69**	1.58***
	SD	1.37	1.27	1.40	1.26	1.58
	P=0.0031					
4.5	N	44	42	42	44	42
	Mean	2.28	1.25***	1.80**	1.22***	0.90***
	SD	1.31	1.26	1.28	1.00	0.89
	P<0.0001					
6.0	N	44	42	42	45	42
	Mean	2.88	1.70***	1.85***	1.84***	1.13***
	SD	1.40	1.64	1.31	1.49	1.19
	P<0.0001					
7.5	N	44	43	42	45	42
	Mean	2.90	1.94***	2.13***	1.67***	1.29***
	SD	1.34	1.41	1.24	1.39	1.34
	P<0.0001					
9.0	N	44	43	41	45	42
	Mean	2.82	2.04***	2.35**	1.79***	1.46***
	SD	1.16	1.46	1.41	1.45	1.33
	P=0.0001					
10.5	N	42	42	40	45	41
	Mean	2.66	2.17*	2.44	2.15**	1.45***
	SD	1.31	1.42	1.52	1.69	1.51
	P=0.0047					
12.0	N	42	42	40	44	41
	Mean	1.99	1.91	2.15	1.73	1.59**
	SD	1.25	1.45	1.32	1.26	1.57
	P=0.38					

Note : # : p-value was based on ANOVA analysis by session for treatment difference
(treatment was the only factor)

* p<0.05 : compared with placebo using mixed-effect ANOVA

** p<0.01 : compared with placebo using mixed-effect ANOVA

*** p<0.001 : compared with placebo using mixed-effect ANOVA

III.6 Reviewer's Evaluation and Comments

In general, the rationale of the study design was not clearly described in the protocol. The validity of using Latin squares design with each patient receiving different treatment at different period without repetition to support the time course claim warrants further investigation. Also, the issues related to the validity of cross-over design, such as carry-over effect, whether the wash-out period is adequate, had not been discussed and justified.

In addition, the primary efficacy endpoint and analysis were not clear in the original protocol or SAP. The concerns of this reviewer are :

1. The treatment action overtime had not been clearly defined. The sponsor only indicated that the analysis would be conducted **by each session**, but did not mention how to define the time to treatment onset or time to loss of effect in the SAP.
2. The detailed description of the mixed effect model had not been provided in the SAP nor reported in the NDA submission. There were no justification or references that shows the validity of the model for this type of design.
3. The impact of missing data was more complicated in this trial. Missing data would occur not only by missing treatment period, but also by missing measurement within a time point. The impact of the missing time point within a period on the analysis and interpretation of the data is not clear.
4. The sponsor did not address multiplicity issues related to multiple endpoints (SKAMP attention and deportment scores) and multiple testing over 8 sessions.
5. In addition, PERMT (attempted, correct) was analyzed as if it was a co-primary endpoint in the study report. In fact, based on the protocol, it should be a secondary efficacy endpoint.
6. The primary hypothesis had never been specified in the original protocol or SAP. The sponsor never mentioned how to assess the overall treatment effect of this study.
Note : Only in the final study report, the sponsor indicated that the overall treatment effect (**average across the 8 scores observed on the classroom day**) was tested first for significance.
7. In addition, the time course was originally proposed as the secondary efficacy analysis.
course. However, it was presented as the primary efficacy section in the report.

This reviewer confirmed the sponsor's results of significant treatment effect **averaged across sessions** in favor of SLI381 doses and Adderall 10 mg, based on SAMP attention and deportment scores. However, due to the earlier concerns, this reviewer did not agree with the sponsor's analysis and interpretation of the time course. Particularly, it was not

clear whether any pair-wise comparison at each time point can be performed if the overall treatment effect at each time point was not significant (see p-values in the second column in Tables II.A.3 and II.A.4).

IV. SUMMARY AND CONCLUSION

In study 381.301, the sponsor had demonstrated significant treatment effect of 10 mg, 20 mg and 30 mg SLI381 versus placebo based on the primary endpoint : average CGIST total score at the last treatment week. The same comparison for the morning and afternoon average CGIST total score at the last treatment week also appeared to be statistically significant. There were concerns about that the sponsor did not clearly indicated their intention to include the morning and the afternoon average CGIST total scores for the labeling claim in their protocol or SAP and the statistical decision rule was not specified. However, based on the conditional testing rule, the treatment effect during the overall, morning and afternoon sessions appears to achieve statistical significance. In addition, after the Dunnett's adjustment, all pair-wise comparisons between each dose level and placebo also appear to achieve statistical significance.

With regard to by-week analysis of the mean CGIST total score, this reviewer found that the 10 mg SLI381 was effective as early as week 1 for overall, morning/afternoon evaluation. Note that due to the forced titration scheme, only 10 mg SLI381 was administered at week 1.

In study 381.201, this reviewer found that the treatment effect (all SLI381 dose levels) was highly significant based on the averaged SKAMP (attention or deportment) scores over 8 treatment sessions. However, due to the issues indicated in section III.6, it is difficult to use such data for efficacy claim about time course.

Yuan-Li Shen, Dr. PH
Mathematical Statistician

Concur :

Dr. Jin _____

Dr. Chi _____

**This is a representation of an electronic record that was signed electronically and
this page is the manifestation of the electronic signature.**

/s/

Yuan-li Shen
7/26/01 09:12:00 AM
BIOMETRICS

Kun Jin
7/26/01 09:35:48 AM
BIOMETRICS

George Chi
7/26/01 10:23:35 AM
BIOMETRICS